DOCUMENT RESUME

ED 043 677 TM 000 146

AUTHOR	Hawkridge, David G.; Campeau, Peggie L.
TITLE	Developing a Guide for Authors of Evaluation Reports of Educational Programs. Final Report.
INSTITUTION	American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY	Bureau of Elementary and Secondary Education (DHEW/OE), Washington, D.C.
REPORT NO	AIR-855-10-69-FR
PUB DATE	Oct 69
CONTRACT	OEC-0-9-089013-2471 (010)
NOTE	31p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.65
DESCRIPTORS	*Administrator Guides, Authors, Descriptive Writing, Educational Programs, Federal Programs, *Guides, *Program Descriptions, *Program Evaluation, Reports, Research Methodology, Technical Reports, *Technical Writing

ABSTRACT

A complete description of the development of a guide to help authors improve the quality of descriptive evaluation reports of federally funded programs is provided. Improvements in reports would be particularly useful to planners who must decide if and how programs should be modified. The methods and procedures employed in the development of the guide from the initial version through field tests, which employed local program evaluators, to subsequent revisions are discussed in detail. The field test samples, embracing a wide variety of programs, locales, administrative units, and expertise, are described and the participants listed. Recommendations for the use of the guide are included and it is suggested that a complementary planning guide is needed. (PR)

T,M,E

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

OE/BESE
Title III
Project 089013

AIR-855-10/69-FR

ED043677

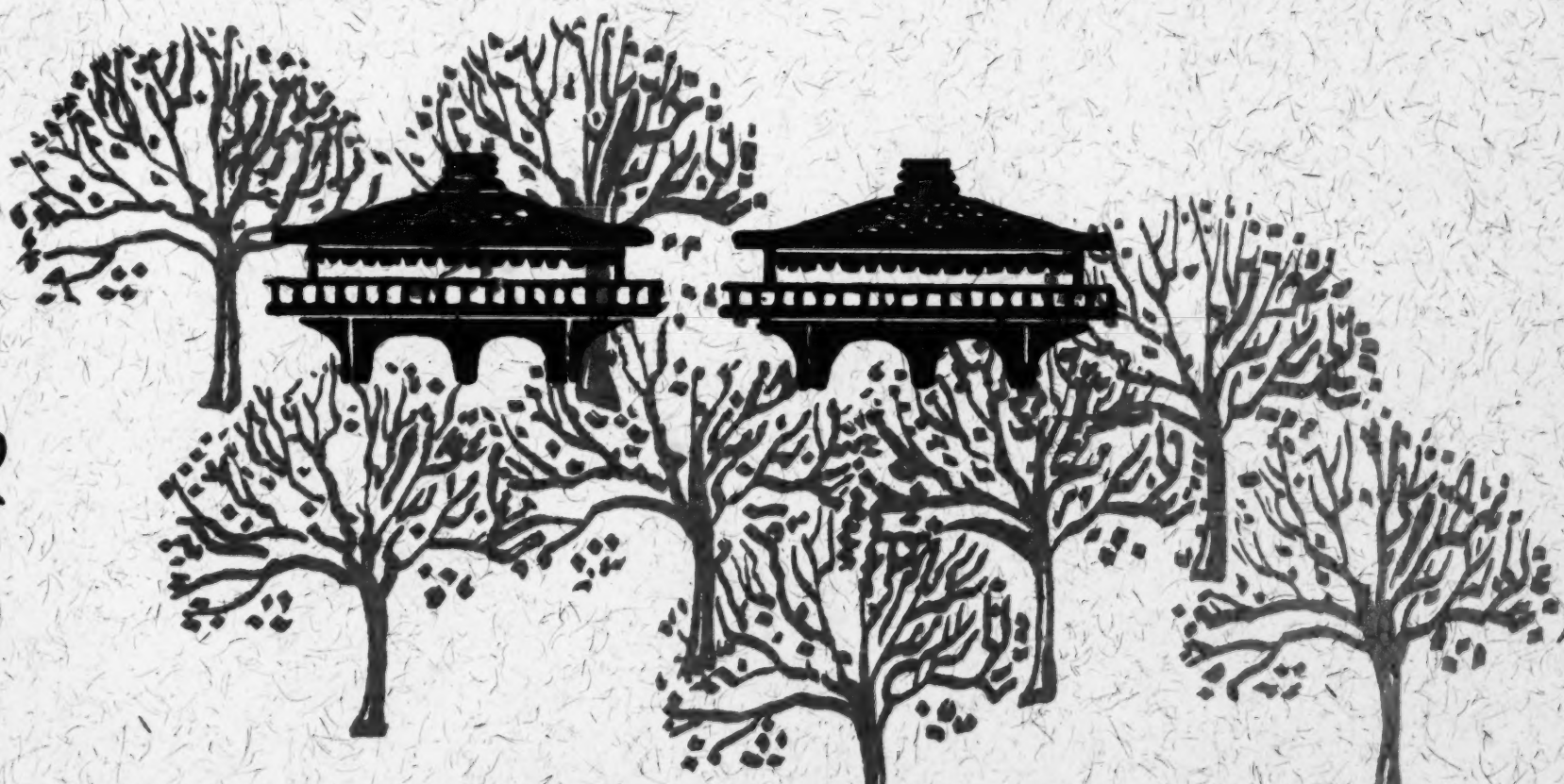
FINAL REPORT

Project No. 089013

Contract No. OEC-0-9-089013-2471 (010) Modification #2

DEVELOPING A GUIDE FOR AUTHORS OF EVALUATION REPORTS OF EDUCATIONAL PROGRAMS

October 1969



AMERICAN INSTITUTES FOR RESEARCH
Post Office Box 1113 / Palo Alto, California 94302

TM 000 146

41

OE/BESE
Title III

AIR-855-10/69-FR

ED043677

FINAL REPORT

Project No. 089013

Contract No. OEC-0-9-089013-2471 (010) Modification #2

DEVELOPING A GUIDE FOR AUTHORS OF EVALUATION REPORTS OF EDUCATIONAL PROGRAMS

David G. Hawkridge

Peggie L. Campeau

American Institutes for Research
in the Behavioral Sciences

Palo Alto, California

October 1969

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education

Bureau of Elementary and Secondary Education

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Background	1
Scope of the Guide	1
Contents of the guide	1
Format of the guide	2
Necessity for Field Testing the Guide	3
METHODS AND PROCEDURES	4
Sampling Procedures	4
Description of Field Test 1 Sample	4
Description of Field Test 2 Sample	5
Field Test 1	6
Revision 1	7
Purpose of the guide	7
Experimental-control-group *paradigm*	8
Process evaluations	8
Jargon and style	8
Checklists	9
Variety and number of sample narratives	9
ERIC	9
Planning	9
Opinion questions	9
Context section	10
Program Description section	10
Evidence section	11
Recommendations section	11
Field Test 2	11
Revision 2	12
Summary section	12
Context section	12
Program Description section	13
Evaluation section	14
Recommendations section	15
References section	16
ERIC	16
CONCLUSIONS AND RECOMMENDATIONS	17
The Effectiveness of the Guide	17
Adoption of the Guide	18
Additional Considerations in Implementing the Guide	18
The Need for a Better Incentive System	19
The Need for Training in Program Planning and Evaluation	20
Summary	21

Page

REFERENCES..... 23

APPENDIX..... 24

Table 1: Major Participants in Field Test 1..... 24

Table 2: Major Participants in Field Test 2..... 26

INTRODUCTION

Background

Early in 1969 the American Institutes for Research in the Behavioral Sciences, under contract to the U.S. Office of Education, prepared a guide to help authors improve the quality of descriptive evaluation reports of Federally funded programs. The present report describes the development of that guide from the initial version through the subsequent revisions which resulted from field testing the guide with samples of its prospective audience.

The need for the guide became particularly apparent during AIR's experience reviewing reports and evaluating programs under earlier AIR-USOE contracts (Hawkridge, Chalupsky, and Roberts, 1968; Hawkridge, Campeau, DeWitt, and Trickett, 1969). Under these contracts AIR surveyed over 1,400 evaluation reports in an effort to identify successful programs. The surveys revealed that many authors of evaluation reports omit critical information about their programs, so that it is impossible to determine the extent to which the program succeeded or failed.

Specifically, report writers tend to omit important details of conditions under which programs operated and of the exact treatments provided and results obtained. For the most part, evaluation reports are at present virtually useless to planners who must decide if a program should be modified, and how, or if it should be continued at all. From the reports it is seldom possible to determine trends toward improvement in education or to identify innovations for dissemination, since program evaluators report little reliable evidence of success.

Scope of the Guide

Contents of the guide. The contents of the guide were originally based on AIR's experience in reviewing reports and evaluating programs during the surveys referred to above. Other inputs were made by Office of Education officials during review sessions and by State evaluation officers at the Third Belmont Conference of the Federal-State Task Force on Evaluation in April, 1969.

In deciding what to include in the guide, the AIR staff made assumptions about the audiences for which evaluation reports are written and the uses to which these reports are put. The report writer's immediate audience probably comprises the superintendent and other district officials. However, it was assumed that evaluation reports also need to be written to meet the needs of people at local, State, and Federal levels who make decisions on future funding policies, who identify and disseminate successful innovations, who determine changes which should be implemented in ongoing programs, or who seek promising models to try in other locales. Thus, the guide was aimed at helping authors produce descriptive reports which would be more useful both for assessing the success of programs and for providing models for others to use or adapt.

Under the terms of AIR's contract with the U.S. Office of Education, the guide was for use at the report-writing stage of a program. Although some of the points covered in the guide have obvious implications for planning a program evaluation, the guide was not designed to serve this purpose, nor would it be an adequate source of help for persons setting up a proper evaluation. A planning guide would necessarily have an entirely different emphasis from a reporting guide and would cover many points which would be merely academic considerations for the report writer who is describing a completed evaluation.

Format of the guide. The original version of the guide, the prototype used in the first field test, comprised three main divisions: Context, Treatment, and Evidence. The Context section suggested information which should be included to adequately describe the background of the program, or what things were like before the program was introduced. The Treatment section dealt with describing the program itself, or who did what, and how. The Evidence section discussed how to report proof of changes brought about by the program.

Each of the three sections included questions which were followed by short explanations, examples, or definitions. At the end of a set of questions, sample narratives were presented as models of how the requested information might be provided in an evaluation report. At the end of each of the three main sections, the questions were repeated to serve as a

checklist. For authors who wished further help and stimulation, references to useful books were provided at the end of the guide.

Necessity for Field Testing the Guide

The guide in its original version (submitted to USOE in April, 1969) was a prototype. It had not been revised on the basis of field testing with its prospective audience, that is, local program evaluators.

The question remained whether the guide would help local evaluators to be more explicit in their reporting. Were some portions of the guide unrealistic in terms of the amount of effort it would take for authors to get requested information? Was the guide usable across a wide variety of programs? Could authors untrained in technical report writing benefit from the guide, or was too much expertise expected?

It was the opinion of the AIR team that field testing the guide with a representative sample of local evaluators would yield a consensus of the kinds of changes which would improve the guide. Moreover, field testing would offer an opportunity to involve State and local people so that they would not later feel that the system was being imposed from above.

A contract was awarded to AIR to field test the guide in the period June-October 1969. The next section of this report describes how the field-test sample was selected, how the two cycles of the field test were carried out, and the kinds of revisions which were subsequently made in the guide. At the end of the report conclusions and recommendations are made, based on the field-test and revision process. The revised guide must be regarded as the chief product of the project, however, rather than the contents of this report.

METHODS AND PROCEDURES

Sampling Procedures

Two cycles of testing and revision were carried out. In each cycle, a sample of 12 local evaluators was asked to use the guide and review it critically. Where consensus among participants indicated a weakness in the guide, revisions were made. The two field-test samples were selected, as much as possible, to exhibit the following variations:

1. The States (6 in each cycle) from which field-test participants were drawn represented major geographical areas of the United States.

2. Field-test participants (2 in each State, or 12 in each cycle, or a total of 24 participants) ranged from university-trained research personnel to program directors untrained in evaluation. In general, expertise of participants depended on the size of the district, and efforts were made to include a fair spread of large (urban and suburban) and small (rural) districts.

3. As much as possible, varied types of programs under several titles of legislation were included in each field-test cycle.

Field-test participants were, in almost all cases, identified and initially contacted by State representatives (officials). Options were limited by certain practical constraints, such as vacation schedules of program personnel and reporting deadlines. Ideally, in the first cycle of the field test, participants were to draw up critiques of the guide while using it for report writing. As it turned out, only half of the participants in the first field test actually reviewed the guide during a period when they were preparing program reports. Every effort was made in the second cycle to obtain collaborators who were indeed writing reports and could therefore test the practical application of the guide. In fact, all those who agreed to collaborate expected to be able to use the guide when writing their reports, but subsequent events prevented 6 out of the 12 from doing so.

Description of Field Test 1 Sample

Major participants in the first cycle of the field test are shown in Table 1 in the Appendix. Some of these participants represented large ESEA

Title I programs, e.g., Long Beach and Garden Grove in California, Belleville and East St. Louis in Illinois, and Minneapolis in Minnesota. Other collaborators were involved in multi-district ESEA Title III programs, e.g., Seattle, Washington, and Fort Worth, Texas. Still other participants were from rural or suburban locales such as the ESEA Title III representatives in Minnesota (Montevideo Public Schools), Washington (Shoreline School District No. 412), and the Civil Rights Act Title IV collaborator from the Wilmer-Hutchins Independent School District in Texas. Finally, there were regional ESEA Title III programs included, such as the two in Georgia, serving Rockdale County and the Ninth Congressional District, respectively.

Besides exhibiting a variety of geographical locations, the first field-test sample also included a variety of types of programs. There were large programs with many components; there was a multi-district cultural arts program, an inner-city reading program, both urban and rural media centers and pupil personnel services centers, a central-cities early childhood education project, a rural-based multi-cultural curriculum development project, and an outdoor education project.

As for relative expertise of the participants in evaluation techniques, three were very well trained, three had virtually no training or skills in research methodology, and the remainder ranked somewhere between the two extremes.

Description of Field Test 2 Sample

Major participants in the second cycle of the field test are shown in Table 2 in the Appendix. As in the first cycle of the field test, the sample selected for Field Test 2 exhibited variation in types of locale, character of programs, expertise of participants in evaluation techniques, and size of administrative units served by programs.

For example, some participants were involved in regional Title III programs, e.g., Shippensburg in Pennsylvania, Florence and Orangeburg in South Carolina, and Jefferson County in Colorado. Other collaborators represented single-district programs: Passaic and Vineland in New Jersey

(Title I), Ontario in Oregon (Title I), and Grove City in Ohio (Title III). Still other participants were in charge of multi-district programs such as the Title I program in Cincinnati, Ohio, and the Title III program in Bethlehem, Pennsylvania.

Types of programs included three reading programs and a guidance-remediation-enrichment program (all under Title I); an individualized instruction project, a teen-tutorial program, a science curriculum project, a regional needs assessment project, a cooperative program for shared services, and a leadership development and instructional improvement project (all under Title III); and, finally, two State-funded migrant programs.

The sites for the various programs included predominantly small, rural communities, e.g., Shippensburg in Pennsylvania, Vineland in New Jersey, Florence in South Carolina, Ontario in Oregon, and Jefferson and Fort Morgan in Colorado. There were also suburban areas represented, e.g., Grove City in Ohio and Orangeburg in South Carolina, as well as inner-city locations such as Bethlehem in Pennsylvania, Passaic in New Jersey, and Cincinnati in Ohio.

Expertise of participants ranged from three individuals with extensive university training in research methodology, to six with little or no formal training in evaluation techniques. The remaining three participants ranked between these two extremes.

Field Test 1

A two-man team from AIR met with each participant for one day to explain and discuss the guide. It was emphasized that AIR was evaluating the guide and seeking reactions to its helpfulness to report writers. Participants were assured that their capabilities and their programs were not being assessed, nor were their report-writing efforts going to be examined formally. Participants were then given four weeks for report writing using the guide. Midway through the four-week writing period, AIR contacted each participant to ascertain if progress was being made and if problems were being encountered in using the guide. In addition, AIR staff were available for telephone consultation on the use of the guide whenever authors felt this was necessary.

(Two collaborators made use of this service.) At the end of the four-week period, the AIR teams returned for two-day visits with participants to discuss their experiences in using the guide. Written comments made by each participant in his copy of the guide were analyzed by the AIR project staff, and notes from the site visits and telephone consultations were studied to diagnose weaknesses in the guide. Where participants consistently had difficulty using the guide, where there were clear indications that the guide did not seem to fit a particular type of program, where additions or deletions were consistently suggested, and so on, the necessary revisions were drafted.

Revision 1

The general reaction of participants to the guide was definitely favorable, the consensus being that it would, in fact, upgrade the quality and usefulness of evaluation reports. Frequently, participants volunteered enthusiastic predictions that the guide's long-term effect would be to improve evaluation by stimulating program personnel to work out better research designs when planning future projects. Comments were mainly positive, and many constructive suggestions were made for ways in which the guide might be improved.

Major changes reflected in the first revision of the guide are discussed in the following paragraphs. General revisions applicable to the guide as a whole are dealt with first. Key revisions specific to one of the three major sections of the guide are dealt with last, under the respective titles of these sections.

Purpose of the guide. The original long introduction to the guide was found in the first field-test cycle to have obscured the general purpose of the guide. Several participants missed the point that the guide was to be used to help them write narrative, or descriptive, reports. Instead, some thought that the questions in the guide were to be answered in sequence, as a sort of questionnaire. Others felt that they had to deal with every point in the guide instead of establishing their own priorities based on the particular characteristics of their programs. The AIR project staff

decided that the purpose of the guide had been buried in the Introduction. Accordingly, this introductory material was rewritten, and retitled The Purpose of the Guide.

Experimental-control-group paradigm. Participants were nearly unanimous in recommending that the guide place less emphasis on the experimental model. They felt this model was more appropriate for university-based research than for action research or demonstration projects. In Revision 1 of the guide, the terms "experimental" and "control" groups were eliminated in favor of "program" and "comparison" groups, respectively. Particularly in the Evidence section of the guide, the general focus was adjusted in order to be more appropriate for programs where proof of success could not be based on the classical treatment-versus-no treatment paradigm.

Process evaluations. The original draft of the guide made no mention of process objectives such as "to provide," "to develop," "to demonstrate," and so on. Participants emphasized that in many programs, simply providing specified services, or simply demonstrating that a project can be implemented, fulfills the contract. In Revision 1 the guide was expanded, particularly in the detailed discussion of objectives in the Evidence section, to take into account these types of goals, and to suggest ways in which their attainment could be measured.

Jargon and style. Although attempts were made to avoid jargon and complicated sentence structure in the initial draft of the guide, the field test revealed occasional lapses by the writers into esoteric discussions related to sampling procedures and data analyses. At other points the writers were justly faulted for ambiguous writing, run-on sentences, and use of jargon which was inconsistent with the "plain talk" approach adopted in the guide. For example, a discourse about acceptable probability levels was deemed too technical and was omitted in Revision 1 of the guide. The title of one of the three major sections of the guide was changed from "Treatment" to "Program Descriptions" because participants felt the first term had undesirable connotations.

Checklists. The first field test revealed that, in general, participants did not need to be reminded of who comprised program personnel, what main segments comprised their programs, what services were provided, and so on. These lists of reminders were almost entirely deleted from Revision 1. (If a planning guide were to be developed, however, such checklists would be relevant to designing a new program.)

Variety and number of sample narratives. Participants were unanimous that the sample narratives were extremely helpful in illustrating how to write up the requested information, but thought the narratives lacked variety in terms of types of programs. Several narratives were rewritten to demonstrate how portions of reports on, say, a work-study program, a bilingual program, or a handicapped children's program might look. Other narratives were adapted to illustrate large, multi-district and small, rural-based programs.

ERIC. The original draft of the guide briefly noted that it would be a good idea for authors to include a completed ERIC resume in their evaluation reports. Most field-test participants were not familiar with the ERIC form and nearly all expressed a desire for instructions on how to complete it. Three pages were added at the end of the guide in Revision 1 to explain how to fill in the various line items. The ERIC Report Resume form was also reproduced and included in the guide.

Planning. Participants pointed out several places in the guide where questions or discussion dealt with matters that amounted to planning considerations, matters which should have been resolved when the program was designed. Unless the requested information was felt to be essential to adequate description of a program, these portions of the guide were deleted. For example, some questions and discussions relating to selection of matched samples were omitted because these considerations were more appropriate during planning phases of the program than to program evaluation.

Opinion questions. To provide authors the chance to express their intuitive conclusions about various aspects of the programs they were describing, the guide contained a few questions such as, "What characteristics

of pupils' activities do you feel contributed to or detracted from the success of the program?" The consensus was that such questions should be dropped, and that conclusions presented in evaluation reports should be supported by evidence presented in the same report. Consequently, the AIR team decided to omit all opinion questions from the revised guide.

Context section. A major overhaul of the Context section of the guide was required to meet the strenuous objections of the participants to the amount of detail requested. Until more districts can afford a computer system to house, update, and provide this information in a directly usable form, participants agreed that most programs could not afford personnel time to compile and compute current demographic data. Thus Context was dealt with much more briefly in Revision 1 of the guide. Only the most essential information was requested on the locale of the program (e.g., population and economic patterns), the school system (e.g., organizational characteristics, curricula, and financial status), and special factors such as needs assessment and historical background.

Program Description section. (Called Treatment in the original draft.) The field test revealed a good deal of redundancy in this section. For example, separate sets of questions on instructional personnel, noninstructional personnel, and services were combined. The original sample narratives on various types of personnel were also combined into a single long narrative which covered all program personnel under appropriate subheadings, and described services not included as part of personnel job descriptions. Similar types of revisions in organization were made in the rest of the section by combining questions and expanding sample narratives.

The original draft of the guide did not include a sample narrative illustrating how program activities might be described by authors, "due to the highly detailed and specific nature of such a description." Without exception, participants agreed that such an example would be most valuable, and that the AIR staff should write one. A lengthy sample narrative illustrating such a description was written for Revision 1 of the guide.

Evidence section. The main objection to this section by participants in the first field test was that it seemed exclusively geared to the classical experimental model. The suggestion was strong and unanimous that to help evaluators of service programs or demonstration programs, the Evidence section should be rewritten to fit other kinds of evaluations, too. For example, the original draft of the guide confined its discussion of program objectives to three areas: cognitive, affective, and psychomotor. In Revision 1, the Evidence section opened with a discussion of process evaluation and gave examples of behavioral objectives for process evaluations.

Several participants urged that the guide be revised to include detailed instructions on specific statistical techniques, but the AIR team felt an adequate "statistical cookbook" was clearly beyond the scope of the guide. However, specific references to useful sources were inserted in the text.

Recommendations section. This section represented an addition to the original draft of the guide. The original draft had concluded the Evidence section with a simple question, "What recommendations can be based upon the conclusions?" Participants felt the guide should give more advice on developing recommendations. The consensus was that too often program evaluators fail to make strong recommendations regarding the future of their programs, even though they often have the most complete evidence upon which to base such recommendations. In the revised guide, an attempt was made to help authors state what should happen next, based on the evidence, and to display their recommendations effectively.

Field Test 2

Participants in the second cycle of the field test are shown in Table 2 in the Appendix of this report. The revised guide which resulted from the first field test was mailed, with only a short covering letter, to the new sample of authors. This time, authors were asked to use the guide for unsupervised report writing; that is, without consulting AIR on the use of the guide. After a four-week period, drafts of their reports were examined as a basis for final revision of the guide. AIR teams then spent one day

with each participant to discuss how the guide might be improved still further. The revisions which were made on the basis of the second field test are summarized below.

Revision 2

Participants' general reaction to the guide was favorable and in several instances enthusiastic. Again, the comment was frequently volunteered that the guide might upgrade program planning in the long run, since some items raised points which should have been taken into account during proposal development or program operation. It was also the consensus of Field Test 2 participants that the organization of information requested in the guide was logical and clear, and that the selection of items under the various headings represented the most essential and relevant information about a program. This general endorsement of the guide's emphasis was in marked contrast to the reaction of Field Test 1 participants to the original version of the guide. Field Test 2 participants predicted that reports prepared along the lines suggested by the guide would be much better sources of important program information than the reports produced under currently available guidelines.

In the first cycle of field testing, most major criticisms by participants applied to the guide as a whole. In the second cycle, this was not so. Therefore, the major revisions suggested are discussed below under the main section of the guide to which they refer.

Summary section. Participants advised that more emphasis be placed on summarizing the contents of evaluation reports. A Summary section was written, and this new addition became the guide's first section. In it, techniques of writing short but adequate summaries were described. A suitably brief, comprehensive summary of a hypothetical evaluation report was prepared and included as a sample narrative.

Context section.

1. Many participants felt they probably would not take the time to write up context information in narrative form. They preferred to make

project applications available to readers interested in this sort of information. The final version of the Context section includes the suggestion that in these cases, suitably edited context statistics and needs-assessment data in the project application be reproduced and included in the report. The author could then refer the reader to this excerpted material after a few sentences which at least highlighted key aspects of the setting for the program.

2. One participant suggested the addition of a question on how priorities among needs were established, once needs had been assessed. The point here was that typically there are more needs than can be met by any one program. A question was added to the Needs Assessment items to identify the basis for selecting needs to be met by the program. Also, a sample narrative was written to illustrate an instance of establishing local priorities from a set of educational needs identified by a State department of education.

3. Multi-district program personnel indicated they would appreciate additional examples of how to describe the diversity of their districts. Alternatives were suggested in the revised version of the Context section.

Program Description section.

1. Participants cautioned that, in spite of the freedom of choice offered in the guide, many report writers would follow the guide's sequence of items. In such cases, "program objectives" and "program participants" would not be identified until the evaluation section of the report. To encourage report writers to briefly specify this information as a preliminary to detailed descriptions of program activities, organizational matters, personnel qualifications and duties, and so on, a new section called Scope of the Program was added at the beginning of Program Description. Under Scope of the Program, the report writer is asked to summarize major program objectives and numbers and kinds of persons served by the program before going into details of the program itself.

2. Participants representing service programs and programs involving teachers or others besides pupils suggested that questions be changed or

added under Activities to make the section more relevant to their programs. By rewording several questions, it was possible to accomplish this and to take into account programs directed at others besides children.

3. The items covering Instructional Equipment and Materials were expanded in the final version of the Program Description section to account for purchased materials which had been adapted to fit a particular program. Participants felt that modifying on-the-market materials was at least as typical a practice as developing original special materials for programs dealing with special groups.

4. Some participants wished to see a less directive approach in the guide to the reporting of Parent-Community Involvement, as the guide's questions under this heading explored politically sensitive areas. It was made clear in the final version of this section that there was no need to describe efforts to enlist parent involvement except where these efforts were an integral part of the program.

Evaluation section. (Called Evidence in the earlier versions of the guide.)

1. Inevitably participants brought up their evaluation problems during site visits from the AIR team. For programs with objectives like "to provide," "to demonstrate," "to develop," and "to increase (specified services)," claims of success were usually based on the fact that the program had been implemented. Participants representing such "process evaluation" felt the Evaluation section should devote more space to the sorts of evidence they could analyze for indications of the quality of the program. In the final version of the guide, discussion of process evaluation was given more emphasis, not only in the introductory discussion at the beginning of the Evaluation section, but also in explanations and examples under Objectives and Analyzing Data. Stress was placed on the type of evidence which provides a basis for conclusions and recommendations about specific modifications which would improve the program.

2. Many participants saw "research" and "evaluation" as separate activities for which different people were responsible. In the words of one participant, the evaluator does research to see if the measurable objectives were met. The measurable objectives may not be the same as the specified "goals" of the project; in fact, they may have been the evaluator's translation of ambiguously worded goals. According to the same participant, the program director, who sits in a different chair and wears a different hat, "evaluates" the program in terms of whether its broader goals were met. The director might claim or be convinced that the program was a success. On the other hand, his evaluator might conclude that, based on objective evidence, a given set of selected measurable objectives was not met. The point, participants agreed, is that the director and the evaluator must resolve their conflicting conclusions about the program. The discussion under Objectives, Measuring Changes, and Reporting Findings was expanded to deal with reconciling judgements of program directors and program evaluators.

3. Participants expressed the need for additional guidance on how to evaluate programs which do not involve pupils and do not seek to improve academic achievement of pupils, e.g., teacher training programs, curriculum development projects, pupil personnel services centers, and leadership development projects. The Evaluation section of the guide was revised at appropriate points to include examples of evidence which could be used to support claims of success for these types of programs.

4. Finally, some participants made the wry observation that sample narratives in the Evaluation section of the guide reported only "successful" programs, with no instances of programs which had achieved little or no success. Three narratives were revised to illustrate programs which failed to achieve at least one objective or which failed to earn the approval of the community.

Recommendations section. The final version of this section was expanded to include some discussion of the situation in which a program director and a program evaluator might disagree on what action should be recommended. (See discussion in the first paragraph under Evaluation on the preceding page.)

References section.

1. Participants frequently suggested titles of additional books or articles to include with the References at the end of the guide. If the suggested sources could be obtained by the AIR team, these materials were reviewed in order to judge their suitability for users of the guide and to assign each acceptable addition to one of the subsections of the References section. (The subsections were as follows: Research Methodology and Experimental Design, Sampling, Measurement--Test Theory, Measurement--Test Construction, Measurement--Periodicals Relevant to Educational Testing, Analyzing Data, and Data Processing.)

2. Participants also frequently commented that it would be helpful if the best or the easiest books could be identified in some way. The AIR team felt there was not sufficient time to study every book in the list of References thoroughly enough to be able to make the necessary comparisons to determine which books were "best." However, every Reference entry was examined in order to assign it a difficulty level: easy, harder, or difficult. An explanatory paragraph at the beginning of the References section was added to explain this classification scheme. Finally, each of the subsections of the References section was reorganized so that all "easy" sources were listed first, "harder" sources were listed second, and "difficult" sources were listed last. It was the feeling of the authors of the guide that the easy-harder-difficult format within each of the subject-matter categories of References had two advantages. First, the use of difficulty levels took into account the wide range of expertise expected in the guide's prospective audience (and perceived during site visits by the AIR team). Second, retaining the broad subject-matter categories facilitated use of the References section by individuals who needed help with a particular aspect of evaluation, such as choosing evaluation samples or analyzing date.

ERIC. When the instructions for completing an ERIC Resume form were cleared with USOE, several changes were suggested. In the final version of the guide, the sample form was omitted, the description of the ERIC system was amplified, and a sample abstract was provided.

CONCLUSIONS AND RECOMMENDATIONS

The Effectiveness of the Guide

The chief aim of field testing and revising the guide was to develop an effective tool for improving evaluation reports. The evidence that shows this aim was achieved is mostly subjective. That is to say, the opinions passed on the guide by all participants in the two successive tryout-and-revision cycles were generally favorable, and it was clear that the guide will meet a felt need if published.

A number of reports were prepared by authors who had the guide at hand, but no rigorous experimental methods could be imposed on those authors by the staff of this project. For example, it was not possible to identify two matched groups of local evaluators, to compare reports produced by a group who had used the guide and a group who had not. Consequently, it is impossible to quantify the effects of the guide. No reports were redrafted using the guide, and no valid comparisons were possible between reports written, say, last year, before the guide was available, and those written this year with the guide.

Although the guide became more universal in its application as revision proceeded, it was clear by the end of the project that the suggestions contained in the guide were most useful to directors of in-the-classroom programs and slightly less useful to directors of service programs. The emphasis in the guide on evaluating the product of a program rather than describing the process did not appeal so much to directors of pupil service centers, for example, as it did to directors of reading programs. There is no evidence available from which to judge the effectiveness of the guide in programs related to dropout prevention, juvenile correction, vocational education, handicapped children, or bilingual education, since such programs were not identified and selected by State departments of education for participation in field testing.

It seems likely that the guide will be of most use in programs such as those under Titles I, III, and VIII of the Elementary and Secondary Education

Act, under Title IV of the Civil Rights Act, and under similar State and local legislation. Programs funded under the Vocational Education Act and other Federal, State, and local legislation could probably use the guide with minor adaptations.

In all cases, the amount and kinds of detail to be reported are considered to be matters for local decision. The guide was not designed nor intended to be an inflexible blueprint designating exactly the information to be included in a report.

Adoption of the Guide

Adoption here does not mean formal declaration by an authority that the guide must be used. Rather, adoption of the guide by, say, a State authority merely indicates the willingness of that authority to encourage use of the guide in the programs where it may be helpful.

The relationships established during the project between AIR, USOE, State departments of education (including the Belmont Federal-State Task Force on Evaluation), and local school personnel were valuable in many ways, but particularly in indicating various routes by which the guide might be adopted. The discussions held between AIR and the other cooperating agencies involved in developing and field testing the guide led the AIR team to the view that State departments of education should introduce the guide, rather than USOE. Funding of programs is increasingly controlled by State departments of education, and it is to them that program directors look for advice and criticism, rather than to the more distant U.S. Office of Education. The State departments of education probably have manpower available to assist in the adoption of the guide; moreover, State personnel know local conditions better than USOE personnel.

Additional Considerations in Implementing the Guide

Should a particular State department of education wish to adopt the guide, merely offering it free of charge to those who request a copy is one possibility for dissemination. A more effective plan might be to make the guide part of an informal system through which authors using the guide could be monitored and from which they could obtain help if required.

Workshops frequently become vehicles for introducing instructional materials and for training persons to use them. In the opinion of the AIR team, a workshop would not be essential to introduce the guide to program evaluators, because the guide was designed to be self-instructional. There is no compelling reason why a training program in report-writing should be needed to supplement the guide.

On the other hand, there is the possibility that the guide might be introduced at the end of a workshop on planning and evaluating programs. Even in that setting, however, potential authors should not be asked to write "instant" reports using the guide, since that would be a very artificial exercise.

The compatibility of the guide with other evaluation instruments was discussed at times during the guide's development. When the guide is introduced to local evaluators, they should be made aware that it was designed to complement instruments used to collect statistical data. Such instruments set up forced-choice categories from which the evaluator or program director has to select the one most nearly appropriate. The guide contains no such compulsions, but suggests topics and details that should be included in a narrative report of a program. The combined reporting power of a narrative report prepared using the guide and statistical reports prepared using the instruments now under development by the Belmont Federal-State Task Force on Evaluation and by the Educational Testing Service should be great indeed.

The Need for a Better Incentive System

The plans for persuading local evaluators to use the guide will vary from State to State. There is no doubt that use of the guide will entail some extra expenditure on evaluation by school districts if they take its suggestions seriously. Not only will more data collection be required, but also the time required for report writing may be longer, too. (The resulting report need not be lengthy, however.) The extra expenditure will be well worthwhile if the resulting reports provide a more reliable basis for general policy-making and for improving programs than do present reports.

Before any State-developed plans to implement the guide can be effective, however, the present incentive system must be altered. Repeatedly during the field tests AIR personnel met report-writers who were collaborating only as a personal favor to some individual or supervisor. Collaborators who did little work with the guide were ones who pointed out that narrative reports were not required of them. In some States, no report was required at all beyond answers to a few simple questions.

The incentive system would be altered most effectively if funding legislation made proper evaluation and the preparation of satisfactory reports more crucial. For example, if a portion of each grant were earmarked for evaluation and reporting, then payment of that portion might be made contingent upon satisfactory evaluation and reporting by the funded agency.

If such a system were impractical, or could only be introduced over a period of years, more immediate and direct incentives might be necessary, such as paying writers out of State-controlled funds for producing improved reports.

To summarize, report writers will have to arrive at the point of seeing that the submission of improved reports is in their own interest, and to their obvious advantage, before those improved reports will become available to USOE and State departments of education. The guide is only a tool. People will have to be persuaded that the product, the report, is worth making. Then they will use the tool, eagerly.

The Need for Training in Program Planning and Evaluation

The beneficial effects of all aids to reporting, such as the guide, can be of no avail when there has been poor program planning and evaluation. The AIR team has encountered many, many programs for which the reporting has been weak. Generally, these are the same programs that were poorly designed in the first place.

The most the AIR guide can do in these instances is to increase the chances that poorly planned programs will be reported better, although over the long term the guide may do something indirectly to improve planning, too.

What is needed, however, by most school districts is help at the planning stage. Too many evaluations are post hoc, verging on frantic attempts to justify another year of funding by pulling together a few pieces of evidence, mostly subjective, about changes the program had induced.

The AIR team that developed the guide is convinced that a planning guide should be developed as a companion to the reporting guide. The first part of the planning guide should comprise suggestions for authors of proposals, dealing with all aspects of proposal writing including budgeting. The second part should be designed to assist program directors in detailed planning of programs once their proposals are funded. Estimating costs, manpower planning, measuring change, data collection and processing, and practical paradigms for evaluation would be included in this section of the guide. The need to integrate evaluation with the overall program strategy would be stressed.

It is not likely that such a guide could be used by itself as a self-teaching device to turn unsophisticated planners into highly skilled and competent program directors. Rather, the planning guide would become the basic text for workshops on program planning and evaluation, and the constant reference thereafter for program directors. It is true that many books exist on educational research; it is also unfortunately true that they are written with the Ph.D. candidate or the university researcher in mind. They seldom serve the needs of the program directors encountered by AIR during this and other recent studies. The proposed planning guide would fill an expressed need, without doubt, and would be a valuable complement to face-to-face training provided by State departments of education or other agencies.

Summary

1. There is much qualitative evidence that the guide will be effective, but little quantitative evidence.

2. The guide is most likely to be useful to directors of in-classroom programs under Titles I, III, and VIII of the ESEA, and Title IV of the Civil Rights Act, or of similar programs.

3. Adoption of the guide should be voluntary.

4. The States should foster such adoption rather than USOE or local districts.

5. Once adopted, the guide should be introduced to local evaluators in some systematic fashion.

6. The guide should generally be regarded as self-instructional.

7. The guide will not be effective as a tool until writers have greater incentive to prepare improved reports.

8. The guide cannot meet the urgent need for training in program planning and evaluation.

9. Training in program planning and evaluation would be assisted greatly by a planning guide developed as a companion to the Guide for Authors.

REFERENCES

Hawkridge, D. G., Chalupsky, A. B., and Roberts, A. O. H. A study of selected exemplary programs for the education of disadvantaged children. Parts I and II. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, September 1968. Final report for the U.S. Office of Education, Office of Program Planning and Evaluation, Contract No. OEC-0-8-089013-3515 (010).

Hawkridge, D. G., Campeau, P. L., DeWitt, K. M., and Trickett, P. K. A study of further selected exemplary programs for the education of disadvantaged children. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, June 1969. Final report for the U.S. Office of Education, Office of Program Planning and Evaluation, Contract No. OEC-0-9-089013-2471 (010).

APPENDIX

Table 1

MAJOR PARTICIPANTS IN FIELD TEST 1

Location	Names and Titles of Major Participants	Programs
CALIFORNIA:		
Garden Grove	Mr. Homer Legree Director of Federal Projects Garden Grove School District	ESEA Title I program and State-funded programs
Long Beach	Miss Geraldine R. Grant Assistant Director of Research Long Beach Unified School District	ESEA Title I program and State-funded programs
GEORGIA:		
Cleveland	Dr. Sofia Divinagracia Director, Title III Evaluation White County	ESEA Title III 9th District Educational Services
Conyers	Mr. William D. Osborne Evaluator, Rockdale County Public Schools	ESEA Title III Pupil Personnel Services Project
ILLINOIS:		
Belleville	Mr. George Quackenbos Coordinator of Special Programs Belleville Public Schools	ESEA Title I Reading Program
East St. Louis	Mr. Albert Moore Director of Research and Evaluation East St. Louis Public Schools	ESEA Title I program
MINNESOTA:		
Minneapolis	Dr. Richard W. Faunce, Consultant Research, Development, and Federal Programs Minneapolis Public Schools	All ESEA programs and others
Montevideo	Mr. Glen Shaw, Director Mr. John Lindner, Assistant Director Educational Media Center	ESEA Title III Educational Media Center
TEXAS:		
Fort Worth	Dr. Charles Evans Research Manager Central Cities Project	ESEA Title III Early Childhood Project

Table 1 (cont.)

Location	Names and Titles of Major Participants	Programs
TEXAS: (cont.)		
Hutchins	Mr. Earl Cook Assistant Superintendent of Instruction Wilmer-Hutchins Independent School District	Civil Rights Act Title IV Multi-Cultural Program
WASHINGTON:		
Seattle	Mr. Ken Sjolund and Mr. Jack Kukuk Puget Sound Arts and Sciences Program Seattle Public Schools	ESEA Title III Cultural Arts Program
Shoreline	Mr. Edgar F. Neal Director, Outdoor Education Shoreline School District	ESEA Title III Outdoor Education Program

Table 2

MAJOR PARTICIPANTS IN FIELD TEST 2

Location	Names and Titles of Major Participants	Programs
COLORADO:		
Jefferson County	Mr. Sam Clifton Director of ESEA Title III Jefferson County Public Schools	ESEA Title III Individualized Instruction Project
Fort Morgan	Mr. Delbert D. Schmidt Director of Special Programs Morgan County Public Schools	State-funded Migrant Program
NEW JERSEY:		
Passaic	Mrs. Rita Wekseler, Coordinator ESEA Title I Passaic Public Schools	ESEA Title I Reading Program
Vineland	Mr. Melvin Scott Coordinator of Federal Programs Vineland Public Schools	ESEA Title I Reading Program
OHIO:		
Grove City	Mrs. Sharlene O'Bryan, Director Teen Tutorial Program South West City School District	ESEA Title III Teen Tutorial Program
Cincinnati	Dr. Joseph L. Felix, Associate Director Mr. David Biegen, Associate Mr. Ronald H. Nieman, Associate Division of Program Research and Design Cincinnati Public Schools	ESEA Title I Guidance, Remediation, and Enrichment Program
OREGON:		
Ontario	Mr. Alvin Hicks, Director Mr. Lawrence Larsen, Assistant Director Migrant Projects Ontario School District #8	State-funded Migrant Program
	Mr. Tom Williams, Coordinator Title I Ontario School District #8	ESEA Title I Remedial Reading Program
PENNSYLVANIA:		
Bethlehem	Dr. Stephen Rituper, Jr. Director, Education Center	ESEA Title III Science Curriculum Project

Table 2 (cont.)

Location	Names and Titles of Major Participants	Programs
PENNSYLVANIA: (cont.)		
Shippensburg	Mr. Frank L. Hair, Director Educational Development Center	ESEA Title III Planning Grant (Regional Needs Assessment Program)
SOUTH CAROLINA:		
Florence	Dr. John W. Baucum, Director Mr. Roger Stiles, Assistant Pee Dee Regional Supplementary Education Center	ESEA Title III Multi-District Cooperative for Shared Services
Orangeburg	Dr. R. H. Braswell, Director Mr. Roger W. Webb, Research Coordinator Region II Education Development Staff	ESEA Title III Leadership Development and Instructional Improvement Project